



# Discovery, structural characterization, and functional insights into a novel apiosidase from the GH140 family, isolated from a lignocellulolytic-enriched mangrove microbial community

Marcelo Vizona Liberato · Douglas Antonio Alvaredo Paixao · Geizecler Tomazetto · Didier Ndeh · David N. Bolam · Fabio Marcio Squina

Received: 24 July 2023 / Revised: 8 December 2023 / Accepted: 14 December 2023 / Published online: 27 January 2024  
© The Author(s), under exclusive licence to Springer Nature B.V. 2024

## Abstract

**Objectives** Apiosidases are enzymes that cleave the glycosidic bond between the monosaccharides linked to apiose, a branched chain furanose found in the cell walls of vascular plants and aquatic monocots. There is biotechnological interest in this enzyme group because apiose is the flavor-active compound of grapes, fruit juice, and wine, and the monosaccharide is found to be a plant secondary metabolite with pharmaceutical properties. However, functional and structural studies of this enzyme family are scarce. Recently, a glycoside hydrolase family member GH140 was isolated from *Bacteroides thetaiotaomicron* and identified as an endo-apiosidase.

**Results** The structural characterization and functional identification of a second GH140 family enzyme, termed MmApi, discovered through mangrove soil metagenomic approach, are described. Among the various substrates tested, MmApi exhibited activity on an apiose-containing oligosaccharide derived from the pectic polysaccharide rhamnogalacturonan-II. While the crystallographic model of MmApi was similar to the endo-apiosidase from *Bacteroides thetaiotaomicron*, differences in the shape of the binding sites indicated that MmApi could cleave apioses within oligosaccharides of different compositions.

**Conclusion** This enzyme represents a novel tool for researchers interested in studying the physiology and structure of plant cell walls and developing biocatalytic strategies for drug and flavor production.

Marcelo Vizona Liberato and Douglas Antonio Alvaredo Paixao have contributed equally to this work.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s10529-023-03460-1>.

M. V. Liberato · F. M. Squina (✉)  
Programa de Processos Tecnológicos e Ambientais,  
Universidade de Sorocaba, Sorocaba, Brazil  
e-mail: fabio.squina@gmail.com

M. V. Liberato  
e-mail: marcelovliberato@gmail.com

M. V. Liberato · D. N. Bolam  
Biosciences Institute, Newcastle University,  
Newcastle Upon Tyne, England  
e-mail: david.bolam@newcastle.ac.uk

D. A. A. Paixao  
Brazilian Biorenewables National Laboratory (LNBR),  
Brazilian Center for Research in Energy and Materials  
(CNPEM), Campinas, Brazil  
e-mail: douglas.paixao1@gmail.com

G. Tomazetto  
Department of Pediatrics, University of Pittsburgh School  
of Medicine, UPMC Children's Hospital of Pittsburgh,  
Pittsburgh, United States  
e-mail: geizetomazetto@gmail.com

D. Ndeh  
Division of Plant Sciences, School of Life Sciences,  
University of Dundee, Dundee, Scotland  
e-mail: dndeh001@dundee.ac.uk

**Keywords** Apiose · Apiosidase · Glycoside hydrolase family 140 · Pectin degradation · Rhamnogalacturonan-II

## Introduction

The D-apiose (Api) is a non-standard furanose found in all higher plants and is important for plant growth and development (Matoh et al. 2000; Ishii et al. 2001; Ryden et al. 2003; O'Neill et al. 2004). Api (3-C-(hydroxymethyl)-D-glycerotetrose) is a highly unusual branched-chain monosaccharide with a tertiary alcohol, which is a component of the pectic structure of rhamnogalacturonan-II (RG-II) and apiogalacturonan from aquatic monocots (Pičmanová and Møller 2016). Furthermore, Api is present in a large number of naturally occurring secondary metabolites (glycosides) that are commercially appealing because of their beneficial health properties and their role in flavor release in food and beverage production. The aglycones of these glycosides include phenolics, terpenes, terpenoids, saponins, cyanogenic nitriles, alcohols, and lactones (Pičmanová and Møller 2016).

Api has a range of biological and chemical properties that make it useful for various applications in medicine and food industry: apiose is a sugar component in red wines, which could impact in grape juice and wine industries (Kashyap et al. 2001; Maicas and Mateo 2005; Guerriero et al. 2018). More than 1000 apiosylated compounds have been identified as secondary metabolites from plants, which can be used to develop novel pharmaceuticals with antigenotoxic (Kaur et al. 2009), antiulcerogenic, (Shiraga et al. 1988) and antiviral properties (Kernan et al. 1998). Apiose-containing glycoside precursors, including phenolics, terpenes, saponins, and aliphatic and aromatic alcohols, are associated with several vital physiological functions in plants, such as pollinator attraction, symbiosis establishment, and cell wall construction (Guerriero et al. 2018). The enzymatic breakdown of these glycoside precursors releases flavor compounds that are useful for food and pharmaceutical applications (Guo et al. 1999; Sarry and Gunata 2004; Mastihuba et al. 2019). Nonetheless, only a few apiosidases have been comprehensively characterized biochemically and structurally (Dupin et al. 1992; Gunata et al. 1997; Guo et al. 1999; Ndeh

et al. 2017; Mastihuba et al. 2019; Potocka et al. 2021).

The discovery-based omics of natural lignocellulolytic systems has boosted the discovery of new carbohydrate-active enzymes (Tomazetto et al. 2020; de Figueiredo et al. 2021). However, by providing a global view, these omics approaches offer general insights into lignocellulolytic systems (Franco Cairo et al. 2016; Mandelli et al. 2017; Brenelli et al. 2019; Granja-Travez et al. 2020; Carvalho et al. 2022). To support the development of biocatalytic strategies for the conversion of plant biomass to value-added products, it is necessary to move beyond a discovery-based approach and focus on enzyme characterization, engineering, and optimization initiatives (Santos et al. 2010; Alvarez et al. 2013, 2015; Diogo et al. 2015; Pimentel et al. 2017). In this sense, the abundance of available data represents a great opportunity for the discovery of novel biological functions and contributes to the advancement of knowledge in molecular and cell biology, particularly within biotechnological applications (Franco Cairo et al. 2013; Tramontina et al. 2017; Cairo et al. 2022; Liberato et al. 2022).

In this context, the present study aimed to determine the structural and functional properties of a GH140 Apiosidase (MmApi), isolated from a lignocellulolytic-enriched mangrove microbial community (LignoManG) (Paixao et al. 2021). We solved the three-dimensional molecular structure of MmApi at 2.1 Å, disclosing its folding and substrate binding site, which present similarity to the endo-apiosidase isolated from *Bacteroides thetaiotaomicron*. The biochemical data provided insights into the specificity of the enzyme, contributing to fundamental concepts regarding the structure and function of apiosidases, which can be useful in the future for medical and food industry applications.

## Experimental

### Phylogenetic analysis

For phylogenetic analysis, over 200 sequences belonging to glycoside hydrolase family GH140 were selected from the CAZy server ([www.cazy.org](http://www.cazy.org)), covering major taxonomic groups within the bacterial domain. Accession numbers were used to retrieve

sequences from the GenBank database. The GH140 catalytic module sequences were aligned using Clustal Omega (Sievers et al. 2011) and ambiguously aligned positions were removed using BMGE (Crisuolo and Gribaldo 2010). The phylogenetic tree was constructed using maximum likelihood with 1000 ultrafast bootstraps, as implemented in raxML (Stamatakis 2014).

### Molecular cloning

Previously predicted protein sequence containing the DUF4038 domain from LignoManG data were used for the current investigation (Paixao et al. 2021). The sequence (Mmapi) was amplified by PCR (1350 bp) using the primers F (5'- ATATATGCTAGCCAGACT TACACTGTAAGC-3') and R (5'- ATATATGGATCC TTATGGCTTCAGAAAATTC-3') and the total DNA of LignoManG as a template (Paixao et al. 2021). For catalytic domain construction (Mmapi\_Cat), the Mmapi gene was used as a template for primers F (5'- ATATATGCTAGCCAGACTTACACTGTAAGC - 3') and R (5'- CTCGAGTAAAAAGGGCGTGA TTCTATGAG - 3'). The sequences were inserted into the pJET1.2/blunt vector for plasmid propagation and further cloned into the pET-28a(+) vector using the restriction enzymes NheI and BamHI for Mmapi and NheI and XhoI for Mmapi\_Cat. *Escherichia coli* DH5 $\alpha$  was used for genetic manipulations and maintenance plasmids. *E. coli* BL21 (DE) was used as a host for recombinant protein expression.

### Protein expression and purification

*Escherichia coli* BL21 (DE3) cells were grown in LB medium containing 50  $\mu$ g/mL kanamin at 37 °C until an OD<sub>600</sub> of 1.0. The recombinant cells were induced

with 0.5 mM IPTG for 16 h at 20 °C. The cells were harvested by centrifugation and resuspended in 20 mL/L culture buffer A (100 mM sodium chloride and 25 mM Tris-HCl, pH 8.0), followed by disruption using sonication. The soluble fraction was isolated by centrifugation and the supernatant was loaded onto a TALON resin (BD Biosciences), washed with 10 column volumes of buffer B (100 mM sodium chloride, 5 mM imidazole, and 25 mM Tris-HCl, pH 8.0), and eluted with buffer C (100 mM sodium chloride, 100 mM imidazole, and 25 mM Tris-HCl, pH 7.0). The fraction containing the purified protein was then concentrated and loaded onto a HiLoad 16/600 Superdex 200 column (GE Healthcare) for size-exclusion chromatography. Buffer (100 mM sodium chloride and 25 mM Tris-HCl, pH 8.0) was used for column equilibration and protein elution.

### Enzymatic assay

As shown in Table 1, several natural polysaccharides and synthetic oligosaccharides were used as substrate on enzymatic assays. For natural polysaccharides and oligosaccharides, the reactions were tested in different buffers (Tris-HCl and sodium phosphate 20 mM) in a final volume of 100  $\mu$ L. The substrates were added to a final concentration of 2 mM (or 0.5% considering polysaccharides) and the protein concentration varied from 1  $\mu$ g/mL to 5 mg/mL. The reaction mixtures were incubated at 40 °C from 1 to 16 h. Enzymatic activities were quantitatively determined based on the reducing sugar released from the substrates using the 3,5-dinitrosalicylic acid (DNS) method (Damasio et al. 2017). A volume of 100  $\mu$ L of DNS solution were added to final reactions and, after boiling the reaction for 5 min, the absorbance was measured at 540 nm. The reactions of the oligosaccharides were revealed by thin-layer

**Table 1** Carbohydrates tested as potential substrates in MmApi activity assay

	Substrates
Polysaccharide	CMC, PASC, Avicel, bacterial microcrystalline cellulose, xylan, xyloglucan, arabinan, arabinoxylan, laminarin, lichenan, $\beta$ -glucan, mannan
Oligosaccharide	Cellobiose, cellotetraose, cellohexaose, chitobiose, chitotetraose, xylobiose
PNP-monosaccharides	$\alpha$ -Arabinopyranoside, $\beta$ -arabinopyranoside, $\alpha$ -fucopyranoside, $\beta$ -fucopyranoside, $\alpha$ -galactopyranoside, $\beta$ -galactopyranoside, $\beta$ -glucopyranoside, $\alpha$ -mannopyranoside, $\alpha$ -rhamnopyranoside, $\alpha$ -xylopyranoside, $\beta$ -xylopyranoside

chromatography (TLC). TLC plates were placed in glass tanks containing running buffer (50% *n*-butanol, 25% water, and 25% acetic acid) for 45 min. The plates were dried and embedded in different staining solutions depending on the substrate: 3% (v/v) sulfuric acid, 75% (v/v) ethanol, 0.1% (w/v) orcinol monohydrate, for gluco- and xylooligosaccharides, and 42% *n*-butanol, 33% methanol, 17% ammonia, and 8% water for chitoooligosaccharides. Finally, the plates were dried again and heated to approximately 100 °C for visualization. Reactions with monosaccharides modified with *p*-nitrophenol (pNP) were determined by measuring the absorbance at 405 nm as previously described (Cota et al. 2015). Because no significant activity was detected, the concentration of pNP was not determined.

#### Crystallization, X-ray data collection and processing

Purified MmApi was concentrated to 9.0 mg/mL and subjected to initial crystallization trials with a Honey-Bee 963 robot (Genomic Solutions) using sitting drop vapor-diffusion plates at the Robolab facility (Brazilian Biosciences National Laboratory, CNPEM, Campinas, Brazil). The drops were composed of 0.5 µL protein solution and 0.5 µL reservoir solution from commercial kits (Hampton and Qiagen). Refinement of the crystallization conditions were set up in hanging drop vapor diffusion plates, where the drops contained 2 µL protein solution and 2 µL reservoir solution. All plates were incubated at 291 K.

The crystals that were completely grown were added to cryoprotective solution, consisting of 20% glycerol added to reservoir solution, and flash-cooled in liquid nitrogen. X-ray diffraction data were collected at the MX-2 beamline of the Brazilian Synchrotron Light Laboratory (CNPEM, Campinas, Brazil) using a PILATUS 2M detector (Dectris) with a radiation wavelength of 1.46 Å. Images were indexed using XDS (Kabsch 2010) and scaled using Aimless (Evans 2006). The molecular replacement was performed using a Phaser (McCoy et al. 2007). Coot (Emsley et al. 2010) was used for density fitting, and refinement was performed using PHENIX (Adams et al. 2010). The crystallographic structure of MmApi was deposited under PDB accession code 8T9W.

## Results and discussion

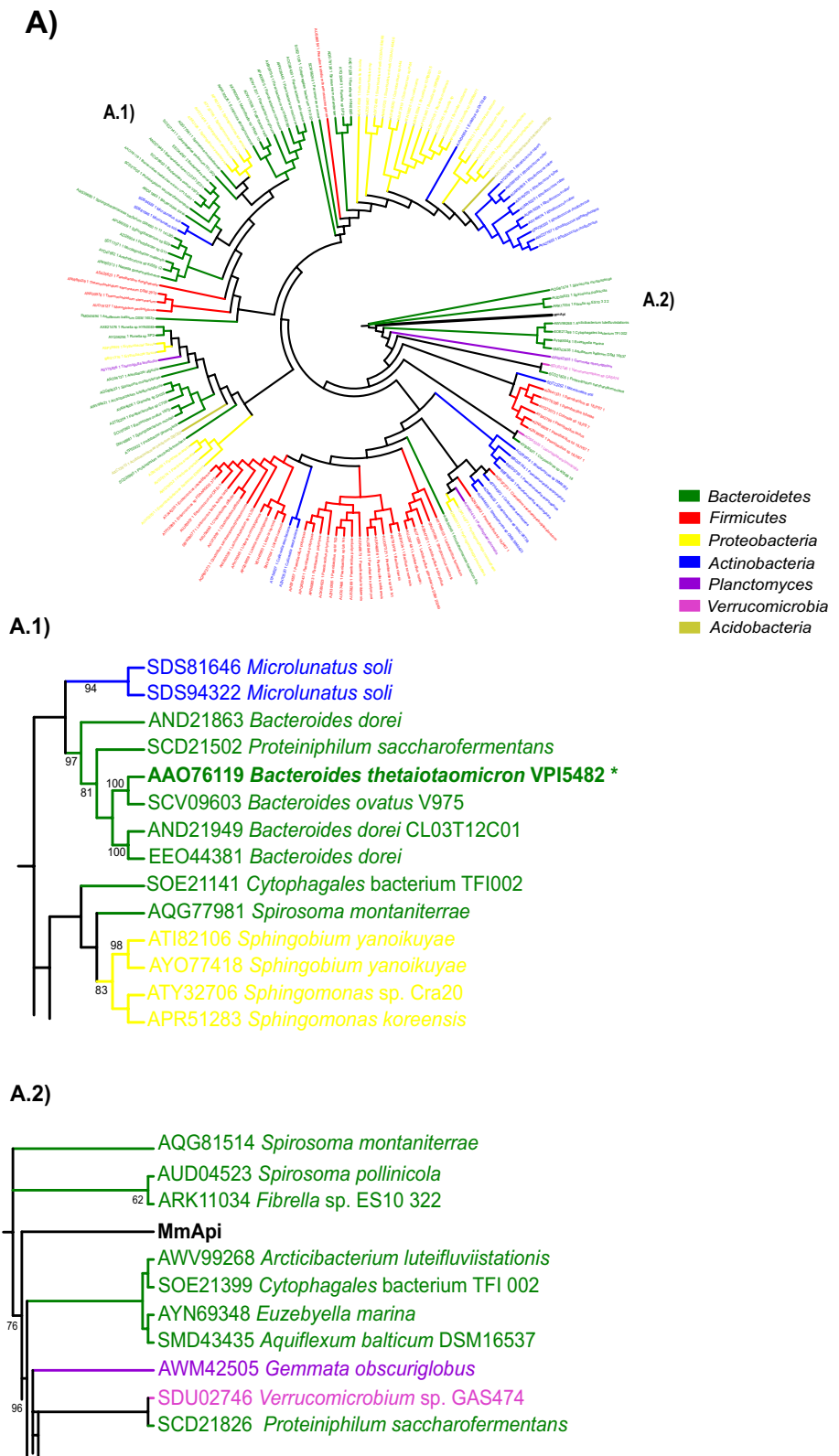
### Gene origin and sequence analysis

In a previous study (Paixao et al. 2021), our group successfully established the LignoMang microbial consortium using a mangrove soil sample and sugarcane bagasse as the microbial and carbon sources, respectively. Similar to the previously described metagenomic discovery platform (Vilela et al. 2023), we analyzed the genetic content responsible for lignocellulose degradation and enzymatic production within the consortium using metagenomic and metaproteomic analyses, along with high throughput screening protocols. The gene encoding the enzyme MmApi was initially identified as a false positive during cellulase screening from a metagenomic library. Consequently, we selected this enzyme for further functional and structural analyses, employing a similar approach as previously described (Campos et al. 2016).

BLASTp sequence search analysis revealed that MmApi shares a high identity with a group of enzymes classified as glycoside hydrolases from family GH140. Functional and structural studies of the GH140 family were first described by Ndeh et al. (2017), where it was referred as BT1012, which was identified as an apiosidase and shares 38.3% identity with MmApi. BT1012 cleaves the  $\beta$ -1,2 linkage between galacturonic acid (GalA), from the RGII backbone, and apiose, and it was the founding member of the CAZy GH140 family.

Based on GH140 sequences retrieved from the CAZy database, most sequences were clustered according to their taxonomic origin at the phylum level, even though located in different clades (Fig. 1A). Notably, MmApi is closely related to members of the *Cyclobacteriaceae* family, specifically the genera *Spirosoma*, *Fibrella*, *Arcticibacterium*, *Aquiflexum*, and the *Flavobacteriaceae* genus *Euzebyella*, all belonging to the Bacteroidetes phylum (Fig. 1B). The sequence exhibits 70% amino acid sequence identity to *Spirosoma montaniterrae* (AQG81514), 69% to *Fibrella* sp. ES10-3-2-2 (ARK11034), followed by 68% to *Spirosoma pollinicola* (AUD04523), 55% to *Cytophagales bacterium* (SOE21399), and *Euzebyella marina* (AYN69348). Thus, we can infer that the MmApi sequence likely originated from the Bacteroidetes phylum. The considerable evolutionary

**Fig. 1** Phylogenetic tree of MmApi based on amino acid sequences, using enzymes classified in the family glycoside hydrolase 140 (<http://www.cazy.org/>). The amino acid sequences were downloaded from the NCBI and only the catalytic domains were used for the phylogenetic analysis. The tree was constructing using maximum likelihood as implemented in raxML. Each protein is labelled by its Genbank accession number, followed the species name. BT1012 (highlighted in bold) represented the first protein sequences from family GH140 with crystallographic structure elucidated and comprehensive biochemical characterization. A.1 and A.2 denote specific branches within the complete phylogenetic tree, encompassing the MmApi and BT1012 enzymes





distance between MmApi and BT1012, both located in different clades, may support divergent enzymatic characteristics enzymes (Aspeborg et al. 2012; Forsberg et al. 2018).

Since MmApi probably belongs to a Gram-negative bacterium, the secretion signal peptide was predicted using Gneg-mPLoc (Shen and Chou 2010). The SignalP server (Almagro Armenteros et al. 2019) designated a cleavage site between residues A21 and Q22. Similar to other GH140 members, MmApi possesses a catalytic domain (residues Q22-F352) and a C-terminal accessory module (AM; L353-P455), which was predicted to be a collagen-binding domain.

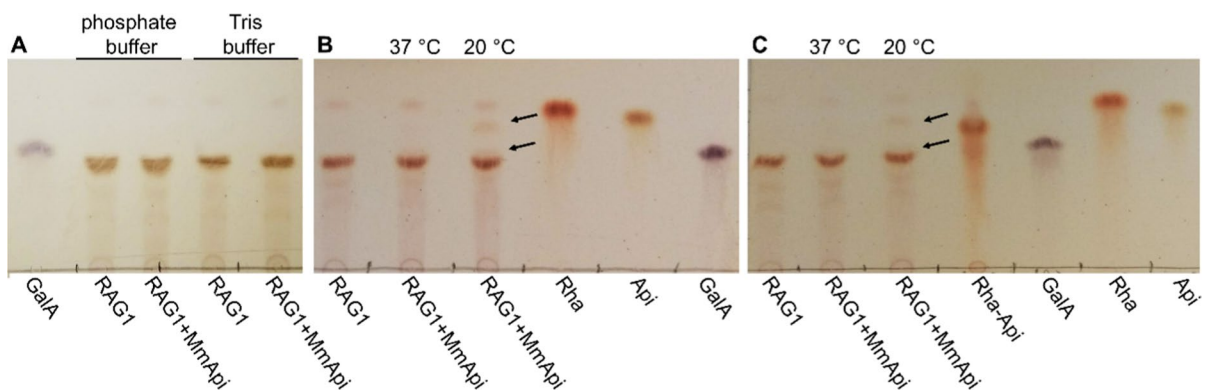
### Catalytic activity evaluation

Two constructs of MmApi were cloned to evaluate the influence of the accessory module on enzyme activity: full-length (MmApi) and catalytic domain (MmApi\_Cat). MmApi was expressed and purified, resulting in 40 mg of protein per liter of cell culture. However, even after testing different *E. coli* strains and expression conditions, MmApi\_Cat was always found to be insoluble (data not shown). This suggests that the C-terminal domain is crucial for the stability of MmApi.

MmApi was subjected to an activity assay using the same substrate prepared for BT1012 (Ndeh et al. 2017). Initially, different concentrations of MmApi (0.1–10 mg/mL) were incubated with 2 mM of the trisaccharide D-galacturonic

acid- $\beta$ 1,2-D-apiose- $\alpha$ 1,3'-L-rhamnose (RAG1), which is a degradation product of RG-II, but no activity was observed after a short incubation period (1 h) (Fig. 2A).

A second assay was performed with overnight incubation at 20 and 37 °C, in phosphate buffer and 5 mg/mL enzyme concentration. After incubation, two bands were observed (indicated by arrows in Fig. 2B), indicating the breakdown of the RAG1 trisaccharide into GalA and D-apiose- $\alpha$ 1,3'-L-rhamnose (Rha-Api). The hydrolysis product resembled that previously observed for BT1012 activity. The degradation products were confirmed through TLC using the disaccharide Rha-Api and GalA as controls (Fig. 2C). According to our data, the control reaction and the MmApi-Cat hydrolysis product exhibit a faint band attributed to the presence of Rhamnose (Rha). In this sense, according to our data, Rha is not a product of the MmApi-Cat activity. On the other hand, it is essential to highlight that although weak, the bands corresponding to the degradation product of MmApi-Cat (GalA and Rha-Api) are only observed as the hydrolysis product. Several reaction conditions were tested (different pH, temperatures, buffers, and the presence of different salts and ions) to determine the optimal conditions for the enzymatic reaction; however, the activity detected was always very low. MmApi was subjected to activity assays using a large range of potential carbohydrate substrates (Table 1) however, no other enzyme activity was detected. Although the MmApi underwent activity assays with



**Fig. 2** TLC analysis of MmApi reactions with RAG1. **A** No activity observed after 1 h of reaction at 37 °C at pH 7.0. GalA was used as reference. **B** Overnight reaction at 20 °C and 37 °C. Rha, Api, and GalA were used as reference. **C** Over-

night reaction at 20 and 37 °C. Rhamnose- $\alpha$ 1,3'-D-Apiose (Rha-Api), Rhamnose (Rha), Apiose (Api), and Galacturonic acid (GalA) were used as reference. The black arrows points to reaction products

a wide range of potential carbohydrate substrates (see Table 1), and various temperatures and buffers were tested, the biochemical data presented in our study offer insights solely into the enzyme's specificity.

### Crystallographic structure

MmApi was crystallized in condition containing 12.5% PEG8000 and 0.1 M HEPES, pH 7.5. The initial phases were obtained by molecular replacement using BT1012 as the template (PDBid 5MSY), and the crystallographic structure was solved to 2.2 Å resolution. The data acquisition and processing statistics are presented in Table 2.

The asymmetric unit contains 36.8% solvent and four protein monomers with high structural similarity among them, calculated as 0.28 Å alpha-carbon root-mean-square deviation (cRMSD) in average. Disregarding the signal peptide for numbering, the monomers were modeled from residues 22 (chains C and D) or 23 (chains A and B) to 449 (chain A) or 455 (chains B, C, and D), respectively. Two residues, K425 and G426, from chain A, located in a loop at the accessory module, could not be modeled due to high mobility.

The final MmApi model clearly shows the presence of two domains that are tightly packed against each other, although we are not completely sure whether such packing occurs in solution or whether it is a crystallographic effect (Fig. 3A). The catalytic domain was comprised of residues Q22–R350 and the accessory module D358–P455, which agreed with the limits previously predicted for MmApi\_Cat cloning.

### Structural features of MmApi accessory module

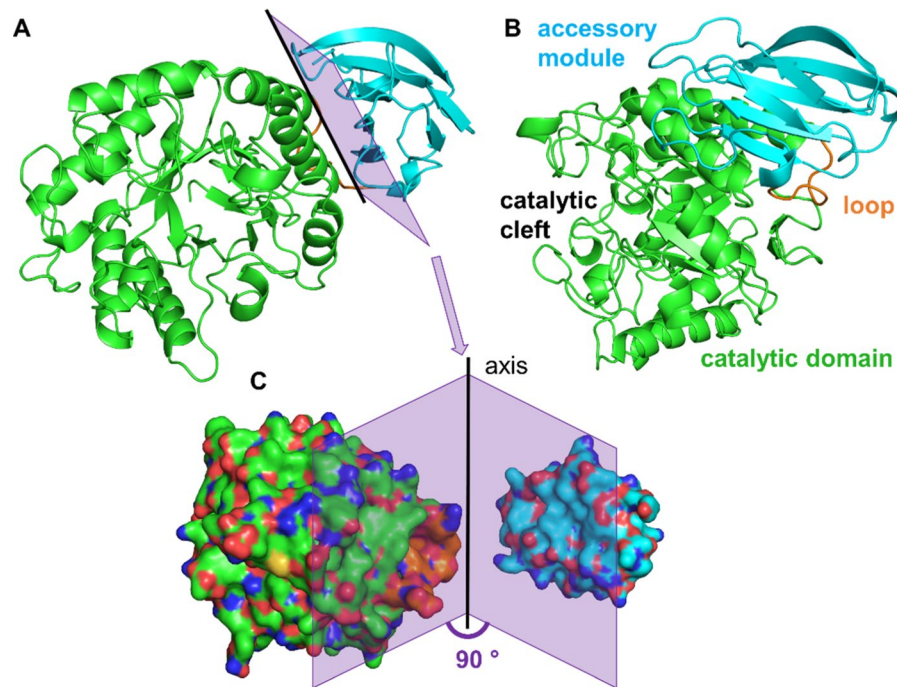
MmApi\_AM is mainly composed of a  $\beta$ -sheet layer with five twisted anti-parallel  $\beta$ -strands, and a second layer of loops, two short  $\beta$ -strands and one short  $\alpha$ -helix (Fig. 3A, B). The most similar structure in the PDB was the AM from BT1012. The subsequent three most similar proteins, which were characterized biochemically and structurally belong to the GH superfamily (Supplementary Table 1). However, all the other three proteins depicted have more than 2.2 cRMSD and reached a maximum sequence identity of 15% in comparison to MmApi (Supplementary Table 1).

**Table 2** Data collection, processing and structure refinement

	Mmapi
<b>Data-collection</b>	
Space group	<i>P</i> 212121
Cell dimension	
a, b, c (Å)	112.09, 113.39, 158.16
$\alpha$ , $\beta$ , $\gamma$ (°)	90, 90, 90
Resolution (Å)	48.19–2.20 (2.24–2.20)
Total number of reflections	655,800 (28,473)
Number of unique reflections	102,431 (5019)
Completeness	0.997 (0.998)
Redundancy	6.4 (5.7)
<i>I</i> / $\sigma$ <i>I</i>	8.5 (1.9)
<i>R</i> <sub>pim</sub>	0.084 (0.486)
CC <sub>1/2</sub>	0.991 (0.704)
<b>Refinement statistics</b>	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.2022/0.2353
No. of non-H atoms	
Protein	13,802
Water	1106
<b>RMSD</b>	
Bonds (Å)	0.003
Angles (°)	0.57
<b>Average B factors (Å<sup>2</sup>)</b>	
Protein	38.12
Water	36.85
<b>Ramachandran plot (%)</b>	
Favoured regions	96.10
Additionally allowed	3.55
Outliers	0.35

Values in parentheses are for the highest resolution shell

As mentioned previously, the AM is packed against the catalytic domain. Several contacts were observed with a clear distribution of buried hydrophobic interactions at the center of the surfaces and polar contacts (hydrogen bonds and salt bridges) on the borders (Fig. 3C). The presence of such hydrophobic areas on the surfaces of both domains indicates that the packing observed in the crystallographic structure may be real in solution. Moreover, the hydrophobic exposed area could be a reason for the instability of MmApi\_Cat observed during the expression trials. Although the AM has a similar length (in number of amino acids) to most CBMs, and a fold similar to the classical CBM  $\beta$ -sandwich, no evidence of a carbohydrate-binding site was identified. AM has no aromatic



**Fig. 3** Crystallographic structure of MmApi. **A** MmApi comprises two distinct domains, which are closely associated and are connected by a short 7-residues loop (orange): a catalytic domain, with  $(\alpha/\beta)_8$  barrel folding (cyan); and an accessory module (mainly composed of a 5-stranded  $\beta$ -sheet and a second layer of loops and short  $\beta$ -strands and  $\alpha$ -helices). **B** 90° rotation of **A** evidencing the catalytic-cleft formed by loops originated from the C-terminal of  $\beta$ -strands. **C** MmApi sur-

face model with the domains rotated 90° apart from each other, where a section (purple area) and an axis (black line) were used as origin. This picture evidences the hydrophobic surface in the center and the hydrophilic surface in the border of both domains at the crystallographic contact interface. Carbon atoms are colored in green for catalytic domain and cyan for accessory module, nitrogen are colored in blue and oxygen in red

residues exposed at the protein surface (Gilbert et al. 2013).

#### MmApi catalytic domain

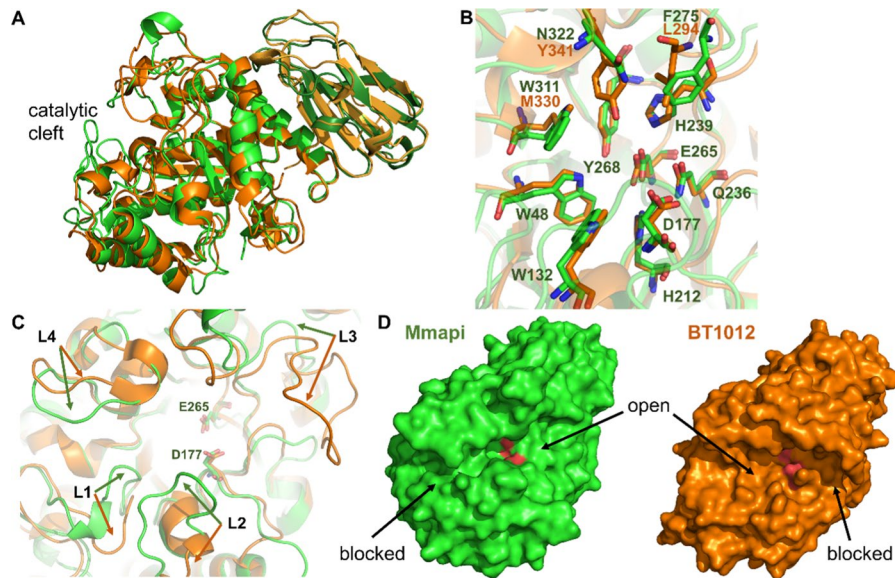
The structure of MmApi\_Cat presents an  $(\alpha/\beta)_8$  barrel, the most widespread fold amongs GH families, where a circular eight stranded  $\beta$ -sheet is surrounded by eight  $\alpha$ -helices (Fig. 3A), with two small N-terminal strands positioned at the bottom of the barrel.

The catalytic cleft is formed by loops that start at the C-terminus of the  $\beta$ -strands and end at the N-terminus of the  $\alpha$ -helices (Fig. 3B). Despite only 38.3% sequence identity, MmApi's model is highly similar to BT1012, with cRMSD of 1.25 Å (Fig. 4A). The catalytic residues, predicted to be D177 and E265, and the surrounding amino acids, which probably coordinate the monosaccharides linked by scissile bond, were identical to BT1012 in composition and

tri-dimensional location (Fig. 4B, C). The exceptions were F275, W311, and N322, which were substituted in BT1012 with the chemically related amino acids L294, M330, and Y341, respectively (Fig. 4B). However, the shapes of the catalytic clefts differ significantly. Loops L1 (connecting  $\beta$ -strand 4 with  $\alpha$ -helix 2) and L2 (connecting  $\beta$ -strand 5 with  $\alpha$ -helix 3) are, respectively, 2 and 7 residues longer in MmApi, and together with loop L4 (connecting  $\beta$ -strand 10 with  $\alpha$ -helix 8), block the binding cleft from one side, while in BT1012 it remains wide open. Conversely, loop L3 (connecting  $\beta$ -strand 8 with  $\alpha$ -helix 6) was 16-residues shorter in MmApi, leaving the other side open, while in BT1012 it is blocked (Fig. 4C).

In summary, these structural characteristics suggest that MmApi and BT1012 can perform the same catalytic mechanism but have different substrate specificities. BT1012 showed the highest activity against GalA-Api-Rha, with a tenfold decrease against





**Fig. 4** Structural comparison between MmApi and BT1012 GH140 enzymes. **A** Superposition of MmApi\_chainC (catalytic domain in green and AM in dark green) with BT1012\_chainA (catalytic domain in orange and AM in light orange) confirms the structural similarity between them. **B** The core of the active-sites are practically identical, leading to the determination of the catalytic residues as D177 and E265 in MmApi. The exceptions are residues whose side-chains are positioned

in the same area and are not chemically distant. **C** The loops L1 to L4 that shape the binding-cleft of MmApi and BT1012 are different in length and position. **D** For MmApi (green), the loops block the binding-cleft at one side (left side in the image) and leave the other wide open (right side in the image). On the other hand, BT1012's loops has an opposite effect, leaving the former side open (left side in the image) and blocking the other (right side in the image)

methylated GalAmet-Api-Rha (Ndeh et al. 2017). Moreover, it is still active with higher polymeric oligosaccharides extending from Rha, which possibly occupy the open side of the catalytic cleft. Interestingly, BT1012 had no activity when arabinofuranose (Araf) was linked to the GalA residue due to steric clash with the blocked side of the catalytic cleft.

It is possible to hypothesize that MmApi activity is restricted to substrates with a lower degree of polymerisation that comes after the apiose monosaccharide due to steric impairment caused by loop L1, and is more permissive to substrates with a higher polymeric degree before apiose. To determine the enzymatic activity of this newly described apiosidase enzyme, alternative approaches, such as high-throughput screening of secondary metabolites, pigments, and plant aromas, or purification of oligosaccharides with varying structures, could be pursued.

MmApi is the second enzyme of the GH140 family to be structurally described and, according to its different features, could potentially be a novel tool for

other researchers interested in studying the physiology and structure of plant cell walls, as well as for the development of drugs and flavors.

**Acknowledgements** We thankfully acknowledge the time provided on the MX-2 Beamline at Brazilian Synchrotron Light Laboratory (CNPEM, Campinas, Brazil) and Robolab at Brazilian Biosciences National Laboratory (CNPEM, Campinas, Brazil). This work was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo by scholarships (2014/04105-4, 2017/12597-2, 2015/23279-6 and 2018/23826-5) and financial support 2020/05784-3 and 2022/08958-8. We are grateful to the National Council for Scientific and Technological Development (CNPq) for financial support 306279/2020-7 and 403844/2023-1.

**Supplementary Information** Supplementary Figure 1—SDS-PAGE of MmApi (52 kDa) expression and affinity chromatography purification.

Supplementary Table 1—Structural and sequential comparison between MmApi and similar proteins described biochemically and structurally.

**Author contributions** MVL: investigation, conceptualization, methodology and writing-review & editing. DAAP: investigation, conceptualization and methodology. GT: investigation,

methodology and writing-review. DN: investigation and methodology. DB: investigation, resources and supervision. FMS: resources and supervision and writing-review & editing.

**Funding** This study was supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (Grant No. 2014/04105-4, 2017/12597-2, 2020/05784-3 and 2022/08958-8), Conselho Nacional de Desenvolvimento Científico e Tecnológico (Grant No. 306279/2020-7 and 403844/2023-1).

**Data availability** The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

## Declarations

**Competing interests** The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

**Ethical approval** No ethics approval is required about this article.

## References

- Adams PD, Afonine PV, Bunkóczi G et al (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr Sect D Biol Crystallogr* 66:213–221
- Almagro Armenteros JJ, Tsirigos KD, Sønderby CK et al (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol* 37:420–423
- Alvarez TM, Paiva JH, Ruiz DM et al (2013) Structure and function of a novel cellulase 5 from sugarcane soil metagenome. *PLoS ONE* 8(12):e83635
- Alvarez TM, Liberato MV, Cairo JPLF et al (2015) A novel member of GH16 family derived from sugarcane soil metagenome. *Appl Biochem Biotechnol* 177:304–317
- Aspeborg H, Coutinho PM, Wang Y et al (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). *BMC Evol Biol* 12:186. <https://doi.org/10.1186/1471-2148-12-186>
- Brenelli LB, Persinoti GF, Cairo J et al (2019) Novel redox-active enzymes for ligninolytic applications revealed from multiomics analyses of *Peniophora* sp. CBMAI 1063, a laccase hyper-producer strain. *Sci Rep* 9:17564
- Cairo J, Mandelli F, Tramontina R et al (2022) Oxidative cleavage of polysaccharides by a termite-derived superoxide dismutase boosts the degradation of biomass by glycoside hydrolases. *Green Chem* 24:4845–4858
- Campos BM, Liberato MV, Alvarez TM et al (2016) A novel carbohydrate-binding module from sugar cane soil metagenome featuring unique structural and carbohydrate affinity properties. *J Biol Chem* 291(45):23734–23743
- Carvalho DB, Paixão DA, Persinoti GF et al (2022) Degradation of sugarcane bagasse by cockroach consortium bacteria. *BioEnergy Res* 15:1144–1156
- Cota J, Corrêa TLR, Damásio ARL et al (2015) Comparative analysis of three hyperthermophilic GH1 and GH3 family members with industrial potential. *New Biotechnol* 32(1):13–20
- Criscuolo A, Gribaldo S (2010) BMGE (block mapping and gathering with entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol* 10:210
- Damasio ARD, Rubio M, Goncalves TA et al (2017) Xyloglucan breakdown by endo-xyloglucanase family 74 from *Aspergillus fumigatus*. *Appl Microbiol Biotechnol* 101:2893–2903
- de Figueiredo FL, de Oliveira ACP, Terrasan CRF et al (2021) Multi-omics analysis provides insights into lignocellulosic biomass degradation by *Laetiporus sulphureus* ATCC 52600. *Biotechnol Biofuels* 14(1):96
- Diogo JA, Hoffmam ZB, Zanphorlin LM et al (2015) Development of a chimeric hemicellulase to enhance the xylose production and thermotolerance. *Enzym Microb Technol* 69:31–37
- Dupin I, Gunata Z, Jc S et al (1992) Production of beta-apiosidase by *Aspergillus niger*—partial-purification, properties, and effect on terpenyl apiosylglucosides from grape. *J Agric Food Chem* 40:1886–1891
- Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr Sect D Biol Crystallogr* 66:486–501
- Evans P (2006) Scaling and assessment of data quality. *Acta Crystallogr Sect D Biol Crystallogr* 62(Pt 1):72–82
- Franco Cairo JPL, Oliveira LC, Uchima CA et al (2013) Deciphering the synergism of endogenous glycoside hydrolase families 1 and 9 from *Coptotermes gestroi*. *Insect Biochem Mol Biol* 43(10):970–981
- Franco Cairo JPL, Carazzolle MF, Leonardo FC et al (2016) Expanding the knowledge on lignocellulolytic and redox enzymes of worker and soldier castes from the lower termite *Coptotermes gestroi*. *Front Microbiol* 7:1518
- Forsberg Z, Bissaro B, Gullesen J et al (2018) Structural determinants of bacterial lytic polysaccharide monoxygenase functionality. *J Bio Chem* 293:1397–1412. <https://doi.org/10.1074/jbc.M117.817130>
- Gilbert HJ, Knox JP, Boraston AB (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol* 23:669–677
- Granja-Travez RS, Persinoti GF, Squina FM, Bugg TDH (2020) Functional genomic analysis of bacterial lignin degraders: diversity in mechanisms of lignin oxidation and metabolism. *Appl Microbiol Biotechnol* 104:3305–3320
- Guerrero G, Berni R, Muñoz-Sanchez JA et al (2018) Production of plant secondary metabolites: examples, tips and suggestions for biotechnologists. *Genes (Basel)* 9(6):309
- Gunata Z, Dugelay I, Vallier MJ et al (1997) Multiple forms of glycosidases in an enzyme preparation from *Aspergillus niger*: partial characterization of a beta-apiosidase. *Enzym Microb Technol* 21:39–44
- Guo WF, Salmon JM, Baumes R et al (1999) Purification and some properties of an *Aspergillus niger* beta-apiosidase from an enzyme preparation hydrolyzing aroma precursors. *J Agric Food Chem* 47:2589–2593

- Holm L, Laakso LM (2016) Dali server update. *Nucleic Acids Res* 44:W351–W355
- Ishii T, Matsunaga T, Hayashi N (2001) Formation of rhamnogalacturonan II-borate dimer in pectin determines cell wall thickness of pumpkin tissue. *Plant Physiol* 4:1698–1705
- Kabsch W (2010) XDS. *Acta Crystallogr Sect D Biol Crystallogr* 66:125–132
- Kashyap DR, Vohra PK, Chopra S, Tewari R (2001) Applications of pectinases in the commercial sector: a review. *Bioresour Technol* 77(3):215–227
- Kaur P, Kaur S, Kumar N et al (2009) Evaluation of antigenotoxic activity of isoliquiritin apioside from *Glycyrrhiza glabra* L. *Toxicol In Vitro* 23:680–686
- Kernan MR, Amarquaye A, Chen JL et al (1998) Antiviral phenylpropanoid glycosides from the medicinal plant *Markhamia lutea*. *J Nat Prod* 61(5):564–570
- Liberato MV, Campos BM, Tomazetto G et al (2022) Unique properties of a *Dictyostelium discoideum* carbohydrate-binding module expand our understanding of CBM–ligand interactions. *J Biol Chem* 298(5):101891
- Maicas S, Mateo JJ (2005) Hydrolysis of terpenyl glycosides in grape juice and other fruit juices: a review. *Appl Microbiol Biotechnol* 67:322–335
- Mandelli F, Couger MB, Paixão DAA et al (2017) Thermal adaptation strategies of the extremophile bacterium *Thermus filiformis* based on multi-omics analysis. *Extremophiles* 21(4):775–788
- Mastihuba V, Potocka EK, Uhliarikova I et al (2019) Reaction mechanism of beta-apiosidase from *Aspergillus aculeatus*. *Food Chem* 274:543–546
- Matoh T, Takasaki M, Kobayashi M, Takabe K (2000) Boron nutrition of cultured tobacco BY-2 cells. III. Characterization of the boron-rhamnogalacturonan II complex in cells acclimated to low levels of boron. *Plant Cell Physiol* 41(3):363–366
- McCoy AJ, Grosse-Kunstleve RW, Adams PD et al (2007) Phaser crystallographic software. *J Appl Crystallogr* 40:658–674
- Ndeh D, Rogowski A, Cartmell A et al (2017) Complex pectin metabolism by gut bacteria reveals novel catalytic functions. *Nature* 544:65–70
- O'Neill MA, Ishii T, Albersheim P, Darvill AG (2004) Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* 55:109–139
- Paixao DAA, Tomazetto G, Sodre VR et al (2021) Microbial enrichment and meta-omics analysis identify CAZymes from mangrove sediments with unique properties. *Enzym Microb Technol* 148:109820
- Pičmanová M, Møller BL (2016) Apiose: one of nature's witty games. *Glycobiology* 26:430–442
- Pimentel AC, Ematsu GCG, Liberato MV et al (2017) Biochemical and biophysical properties of a metagenome-derived GH5 endoglucanase displaying an unconventional domain architecture. *Int J Biol Macromol* 99:384–393
- Potocka EK, Mastihubova M, Mastihuba V (2021) Apiose-relevant glycosidases. *Catalysts* 11(10):1251
- Ryden P, Sugimoto-Shirasu K, Smith AC et al (2003) Tensile properties of Arabidopsis cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II-borate complexes. *Plant Physiol* 132:1033–1040
- Santos CR, Meza AN, Hoffmam ZB et al (2010) Thermal-induced conformational changes in the product release area drive the enzymatic activity of xylanases 10B: crystal structure, conformational stability and functional characterization of the xylanase 10B from *Thermotoga petrophila* RKU-1. *Biochem Biophys Res Commun* 403(2):214–219
- Sarry JE, Gunata Z (2004) Plant and microbial glycoside hydrolases: volatile release from glycosidic aroma precursors. *Food Chem* 87:509–521
- Shen HB, Chou KC (2010) Gneg-mPLoc: a top-down strategy to enhance the quality of predicting subcellular localization of Gram-negative bacterial proteins. *J Theor Biol* 264:326–333
- Shiraga Y, Okano K, Akira T et al (1988) Structures of potent antiulcerogenic compounds from *Cinnamomum cassia*. *Tetrahedron* 44:4703–4711
- Sievers F, Wilm A, Dineen D et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:539
- Stamatakis A (2014) RAXML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313
- Tomazetto G, Pimentel AC, Wibberg D et al (2020) Multi-omic directed discovery of cellulosomes, polysaccharide utilization loci, and lignocellulases from an enriched rumen anaerobic consortium. *Appl Environ Microbiol* 86(18):e00199-20
- Tramontina R, Franco Cairo JPL, Liberato MV et al (2017) The *Coptotermes gestroi* aldo-keto reductase: a multipurpose enzyme for biorefinery applications. *Biotechnol Biofuels* 10:4
- Vilela N, Tomazetto G, Goncalves TA et al (2023) Integrative omics analyses of the ligninolytic *Rhodospiridium fluviatile* LM-2 disclose catabolic pathways for biobased chemical production. *Biotechnol Biofuels* 16(1):5

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.